

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**

[MENU](#)
[SEARCH](#)
[INDEX](#)
[DETAIL](#)
[JAPANESE](#)
[BACK](#)

2 / 2

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 06-290125  
 (43)Date of publication of application : 18.10.1994

(51)Int.Cl. G06F 13/00  
 G06F 12/00

(21)Application number : 06-025562 (71)Applicant : INTERNATL BUSINESS MACH  
 CORP <IBM>

(22)Date of filing : 23.02.1994 (72)Inventor : SHOMLER ROBERT W  
 MCILVAIN JAMES E

(30)Priority

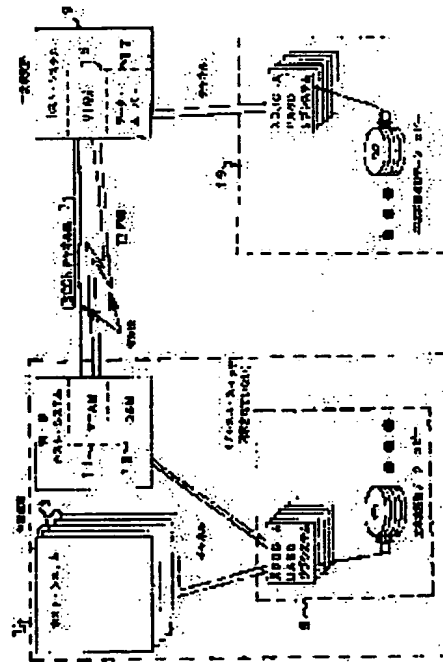
Priority number : 93 36017 Priority date : 23.03.1993 Priority country : US

**(54) METHOD AND SYSTEM FOR REMOTELY BACKING UP AND RESTORING DATA OF MULTIPLEX SYSTEM**

(57)Abstract:

**PURPOSE:** To maintain the consistency of synchronously maintained remote backup copy data in the same state as that for primary-side copy by forming a write token containing a peculiar order number and the address of an external storage device.

**CONSTITUTION:** In response to each writing operation to a DASD external storage device 5, a write token containing a peculiar order number and the address of the storage device 5 is formed. Then a path is set between a secondary working place 9 and the storage device 5 and the list of newly generated write tokens and the relevant updated data of a token list transmitted with the same or a previous message are transmitted as a message selectively containing each message. At the working place 9, a reserved write queue is maintained for inquiring data with the token in response to each message received from a processor and only the coincident tokens and updated data are scheduled and written in the order of the queue.



LEGAL STATUS

[Date of request for examination] 23.02.1994

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

ES978

特開平6-290125

(43)公開日 平成6年(1994)10月18日

(51)Int.Cl. <sup>5</sup>	識別記号	庁内整理番号	FI	技術表示箇所
G 0 6 F 13/00	3 5 1 M	7368-5B		
12/00	5 3 3 J	8944-5B		

審査請求 有 発明の数 5 OL (全 13 頁)

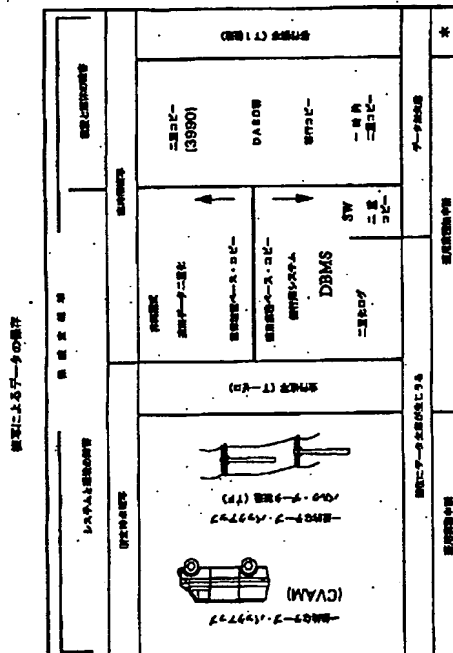
(21)出願番号	特願平6-25562	(71)出願人	390009531 インターナショナル・ビジネス・マシーンズ・コーポレーション INTERNATIONAL BUSINESS MACHINES CORPORATION アメリカ合衆国10504、ニューヨーク州アーモンク (番地なし)
(22)出願日	平成6年(1994)2月23日	(72)発明者	ロバート・ウェズレイ・ショムラー アメリカ合衆国カリフォルニア州、モーガン・ヒル、ビードモント・コート 17015 番地
(31)優先権主張番号	0 3 6 0 1 7	(74)代理人	弁理士 頓宮 孝一 (外1名)
(32)優先日	1993年3月23日		最終頁に続く
(33)優先権主張国	米国 (US)		

(54)【発明の名称】 多重システムの遠隔データバックアップおよび回復を行う方法およびシステム

(57)【要約】

【目的】 本発明は、送信中の更新や、更新コピー発信源が中断された時に受信されなかった更新は除外し、遠隔バックアップコピー・データの安全性を一次側コピーのそれと同じに非同期で維持する方法と手段を提供する。非同期式遠隔データバックアップにより情報処理システムのデータを保存するものであり、そのバックアップは実行中の適用業務を中断せず、さらに伝送中の欠落や主事業所と遠隔事業所間で何らかの中断が発生した時に更新が受信されなかったために生じたデータ欠落が遠隔事業所で解明される。

【構成】 主事業所における各書込オペレーションにตอบสนองしてトークンと固有の順序番号が割り当てられ、このトークンと更新データが遠隔事業所に送信される。この順序番号を用いて順序を確定し欠落した更新データを確認するため順序中にギャップを定める。



## 【特許請求の範囲】

【請求項1】非同期で独立生成された適用業務依存の書込みオペレーションのシーケンスの発生源から遠隔の場所に位置する事業所において、データ・コピー・セットの整合性と回復可能性を与える方法であって、

(a) データ発生源における各書込みオペレーションに  
10 応答して、固有の順序番号とソース・アドレスを含む書込みトークンを形成するステップと、

(b) データ発生源と遠隔事業所間にバスを確立し、各  
メッセージが最近生成された書込みトークンのリスト、  
および同一または以前のメッセージで送信されたトーク  
ン・リストの該当更新データを選択的に含むメッセージ  
を前記遠隔事業所に送信するステップと、

(c) データ発生源から受信したそれぞれのメッセージ  
に  
15 応答し、データをトークンと照合するために保留書込み  
待ち行列を遠隔事業所に維持し、適合トークンと更新  
データについてのみ待ち行列の順で、遠隔事業所で適  
合トークンと更新データをスケジュールしそして書き込む  
ステップと、を含む方法。

【請求項2】前記方法が更に、

(d) シーケンスの少なくともある部分で生じた中断に  
20 応答して、遠隔事業所において維持している保留書込み  
待ち行列を走査し、データの時間的に整合したイメージ  
を発生源における書込み順序と同じ書込み順序で提示  
するように、その事業所で受信した順次トークンおよび  
更新データを解釈するステップ、を含む請求項1の方法。

【請求項3】請求項1のステップ(b)が更に、

データ発生源と複数の遠隔事業所間にバスを確立し、ま  
た少なくとも一つのメッセージを1番目の事業所へ送信  
30 すること、少なくともいま一つのメッセージを2番目の  
事業所へ送信すること、あるいは少なくとも一つのメッ  
セージを1番目と2番目の両事業所へ送信することのい  
ずれかを含め、メッセージを送信するステップ、を含む  
請求項1の方法。

【請求項4】送信中の更新や、更新コピー発生源が中断  
された時に受信しなかった更新を除き、非同期的に維持  
されるデータの遠隔バックアップコピーの整合性が一次  
側のコピーの整合性と同じになるよう保証する方法であ  
って、

40 コピー発生源と遠隔事業所間にメッセージ・インタフ  
ェースを設立するステップと、

発生源における各書込みオペレーションに  
15 応答し、発生源アドレスの表示と固有順序番号を表す書込みトーク  
ンを生成するステップと、

最近生成されたトークン、書込みトークン、および同一  
または以前のメッセージで送信されたトークン・リスト  
の該当更新データをリストしたメッセージを発生源から  
20 遠隔事業所へ送信するステップと、

遠隔事業所において、

各受信メッセージに  
25 応答して、データをトークンと照合

するために保留書込み待ち行列を維持するステップと適  
合するトークンおよび更新データについてのみ待ち行列  
の順で遠隔事業所において適合トークンおよび更新デー  
タをスケジュールしそして書き込むステップと、  
30 中断が発生した場合は、遠隔事業所での更新データ欠落  
の時点を正確に確立するため、書込み待ち行列と生成メ  
ッセージを比較照合するステップと、を含む方法。

【請求項5】プロセッサと、前記プロセッサに常駐  
のオペレーティング・システムと、外部記憶装置と、前  
記オペレーティング・システムへの呼出しによって前記  
外部記憶装置へ更新を書き込むため前記プロセッサで  
40 実行している適用業務に  
15 応答する手段と、適用業務と非同期的に前記外部記憶装置から遠距離に位置する事業所  
に前記更新のバックアップコピーを伝播するための手段  
であって前記更新の前記バックアップコピーを前記遠隔  
事業所で非同期で書き込むための手段を含むものと、を  
有するシステムにおいて、

固有の順序番号と外部記憶装置のアドレスを含む書込み  
トークンを形成するため、外部記憶装置での各書込みオ  
ペレーションに  
20 応答する手段と、

外部記憶装置と遠隔事業所間に通信バスを設定し、その  
間でメッセージを送信する手段であって、各メッセージ  
が、新しく生成された書込みトークンおよび同一または  
以前のメッセージで送信されたトークン・リストの該当  
更新データを選択的に含むものと、

遠隔事業所にあり、各受信メッセージに  
25 応答して、データをトークンと照合するために保留書込み待ち行列を維持し、また合致したトークンおよび更新データについてのみ待ち行列の順で遠隔事業所において、合致したトークンと更新データをスケジュールし書き込むための手段と、

を含む多重システムの遠隔データバックアップおよび回復を行うシステム。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、非同期式遠隔データバックアップ（遠隔バックアップ複写とも呼ばれる）による情報処理システムのデータ保存に関するものであり、特に主事業所の記憶サブシステムに本拠を置くコピーから遠隔地でデータの複写を行うこと、そしてその複写は、適用業務の実行を中断せず、さらに伝送中の欠落や主事業所と遠隔事業所間で何らかの中断が発生した時に更新が受信されなかったために生じたデータ欠落（曖昧性）の理由を遠隔事業所側で解明できるようなデータ複写に関する。このような中断は典型的に自然災害や人為的災害によって引き起こされ、主事業所での作業とデータを利用不可能にし、作業の継続には遠隔事業所で複写され記憶されているデータと対話することが必須となる。

【0002】

【従来の技術】データ複写は情報処理システムあるいは計算機システムにおけるデータ保存の一形態である。しかし、データ複写によるデータ保存には多くの要因を考慮に入れなければならない。主事業所での作業やデータが利用不可能になった際に、遠隔事業所で複写され記憶されているデータがそのデータとの継続対話のためのリポジトリとなることが予測される場合には、このことは特に重要である。複写することについて熟慮すべき要因には、保護定義域（システムや環境の障害、または装置や媒体の障害）と、データ欠落（無欠落、一部欠落）と、他のデータや処理の発生に関連して複写が発生する時間（時点、実時間）と、前記計算機で実行中の適用業務に対する中断の程度と、そして、その複写は適用業務関係のものかあるいは記憶サブシステム関係のものかということが含まれる。前記適用業務関係の複写にはログ・ファイル、データ・ファイル、プログラム・ルーチンを含み、一方前記記憶サブシステム関係の複写には直接アクセス記憶装置（DASD）アドレスの認識とデータ・セット識別子を含む。

【0003】次に説明するように、従来の技術には、回復と保存を目的とするデータのバックアップやデータをバックアップするための方法と手段が多く見られる。オリジナル・データ・セットと複写データ・セットとの多少の不整合は許容できるかもしれないが、このような曖昧さを容認できないものにするのは欠落データの発生の看過である。換言すれば、コピー・セットの曖昧さは、回復プロセスがデータ・コピーの状態を判定できないことを意味している。

【0004】整合性をもつバック・アップ・データ（無欠落複写）を生成するための従来の技術による一方法は、DASDに記憶されているデータの一部を1日1回磁気テープに複写し（一時点複写）、そのテープ媒体をトラックで遠隔地へ輸送する方法である。このように、複写データ・テープのトラック輸送は（1テープ・リール当たり200メガバイトのデータを記憶しており、このテープ200リールを主事業所から約80キロメートル離れた遠隔事業所へ1時間でトラック輸送すると仮定する）平均速度40、000MB/3、600秒すなわち11.111メガバイト/秒で通信していることになる。1950年代頃のこのようなデータバックアップの実施は適用業務を完全に中断してしまうものであった。さらに、その実施は不向きであり、また主事業所と遠隔事業所との間でデータ状態に1日分の差を生じさせる。

【0005】さらに他の方法は、バックアップ・コピーを移動するためにT1回線のような通信回線を使用する方法である。より現在時間に近い（分単位の時間までも含まれる）最新の遠隔複写が、IMS、DB2のようなデータベース管理システムによって提供される。このような遠隔複写はリアルタイムで行われるが、それは両地

点を結ぶ常時使用可能な専用回線を介して伝送されるので、記憶装置関係のコピーよりはむしろ適用業務関係のコピーに依存している。

【0006】Cheffetz 其他による「ネットワークのためのバックアップ・コンピューター・プログラム」に関する米国特許 5,133,065、1992年7月21日付与、は、それぞれのローカル・ノードがバック・アップされるべきローカル・ファイルのリストを作成し、それを送信する宛先となるファイル・サーバーをもつLANを開示している。このような遠隔生成は、リスト作成およびファイル複写活動をネットワーク・サーバーが開始する場合に生じるトラフィックを減少することを目的とする。議論の余地はあるが、この参照例より以前の技術は中央管理されるファイル選択を教示していた。このことは、結局ローカル・ノードの保安とサーバーの過度使用の妥協をもたらした。これはCheffetz のローカル・ノード生成リストおよびファイル・サーバーへのこのリストの送信によって回避されるものと想定される。

【0007】Cheffetz が負荷と保安とを調和させることに幾分の懸念のある一時点複写の一形態を述べていることは明らかである。しかしながら、データの整合性に対して、すなわちファイル・サーバーによって実際に複写されたデータの曖昧性を説明することに対しての用意がなされていない。

【0008】Mikkelsen による「データの無時間差バックアップ複写を行う方法と手段」に関する1991年10月19日出願、審査継続中の米国特許出願第07/781,044号は、論理アドレスと物理アドレスとの一致を形成するための所要時間だけ実行を保留し、その後スケジュールに従い、あるいはその場の都合によって、記憶サブシステムのデータセットを物理的にバックアップすることによって、CPUの適用業務と並行して、DASD記憶サブシステムに関し、CPUでの一時点の整合性をもつ指定されたデータセットをバックアップ複写することを教示している。Mikkelsen の方法と手段は適用業務を中断せずに選択された記憶装置関係のデータセットの一時点複写を行うのに有利である。それでも、一時点データは同時的複写が開始された時点でのデータの断片として遠隔事業所に到着するという欠点がある。このことは、次に続くデータの更新に比較しデータの状態が何時間も古いものであることを表している。非同期式の遠隔データバックアップは、リアルタイムで行われる複写オペレーションであり、更新は継続的に先送りされるものであることに注意する必要がある。

【0009】Dishon 他による「同期式磁気ディスク駆動装置における多重コピー・データの仕組み」に関する米国特許 4,862,411、1989年10月29日付与、は、チャンネル・コマンド・ワード・チェーンとは無関係に同期している一対のDASDへの並列書込みバスを作ることによってバックアップコピーを確実なものにしており、Croc

kett 他による「非同期で稼動する周辺装置の制御」に関する米国特許 4,837,680、1989年6月6日付与、に示されているような単一の制御装置によるバックアップコピー・バスを回避している。IBM システム/370および同種のシステムのようなシステムにおいては、チャンネルと呼ばれる仮想計算機が、バスの設定と、CPUの主記憶装置と外部記憶装置間のデータ転送とを管理するために使用されていた。制御は、解釈と実行のため外部記憶装置の制御装置へチャンネルによって伝えられるチャンネル・コマンド・ワード (CCW) の形式で与えられていた。

【0010】Dishon はリアルタイムかつ適用業務無中断の複写を例示しており、それは媒体や装置の障害に備えてデータを保存するが、主事業所側での障害が引き起こすようなシステムや環境の障害は扱わない。

【0011】Beale 他による「データの冗長コピーを提供するためのデータ記憶システム」に関する米国特許 5,155,845、1992年10月13日付与、は、第一の記憶制御装置によって書き込み操作を処理させ、直接リンク (広帯域バス) を介して第二の記憶制御装置に並行通信させることによって2台以上の外部記憶装置へ可変長レコード (CKD: カウント・キー・データ方式) バックアップ複写を行う。これは主事業所と遠隔複写場所間のバス長の制限を不要にしている。CKDの要求/応答アーキテクチャーが長さ (帯域幅) を約150メートル程度に制限している事実から、このような制限が生じている。

【0012】Beal はDishonの方法で、装置や媒体の障害が発生した場合のみデータの可用性を保つ、リアルタイムかつ適用業務無中断のバックアップ複写を扱っている。

【0013】別の表現をするなら、非同期でかつ主事業所とは無関係に、遠隔事業所側を更新するために、主事業所でどのようにして書き込み更新操作が生成されるかを認識する必要がある。このことについて、主事業所はプロセッサで並行して実行している一つ以上の適用業務を含んでおり、それぞれの適用業務は「適用業務依存書き込み」と呼ばれるものを生成する。すなわち、記憶サブシステムは、オペレーティング・システムから呼び出されるべき書き込み操作や、その待機スケジュールについての知識や認識をもっていない。遅延の変動は局所的なものである。適用業務は単に1台のあるいは同じDASDへ書き込むだけではない。事実、初期書き込みとコピー生成のために、待ち行列の長さおよび待ち行列のサービス率がともにDASD記憶装置によって変動するように、適用業務は書き込みを異なったパターンに郡別する。これは遠隔事業所で受け取ったコピーが、長時間にわたって順序が揃ったり乱れていたり、また遅延や欠落を起こす確率はかなり高いことを意味している。

【0014】一言で言えば、非同期でかつ独立して実行している適用業務およびプロセッサは、ローカル記憶

装置と遠隔事業所に対して書き込み操作のストリームを作成する。そのストリームは待機させられ、異なる速さで実行されるため、ほとんどランダムに近いコピー順序をもたらす。

【0015】

【発明が解決しようとする課題】本発明の目的は、送信中の更新や、更新コピー発生源が中断された時に受信しなかった更新は別として、非同期的に維持される遠隔バックアップコピー・データの整合性を一次側コピーのそれと同一に維持する方法と手段を提供するにある。

本発明の他の目的は、更新コピー発生源と遠隔事業所間のDASD記憶装置の管理レベルで蓄積送信メッセージ・インターフェースを使用するための方法及び手段であって、中断があった場合は遠隔事業所で更新の完全性についての相違や欠落を完全に特定できるような方法と手段を提供するにある。

【0016】本発明の他の目的は、非同期でかつ独立して生成されたコピーと、主事業所から一ヶ所以上の遠隔データバックアップを行う事業所へ与えられる情報とによって、データ・セットの順序と完全性を回復するための方法と手段を提供するにある。

【0017】

【課題を解決するための手段】本発明の目的は、以下の構成によるシステムで使用される方法と手段によって満たされる。すなわち少なくとも1台のプロセッサ、少なくとも1インスタンスの前記プロセッサに常駐のオペレーティング・システム (OS)、DASD外部記憶装置、前記OSへの呼び出しを通して外部記憶装置へ非同期でかつ独立して更新を書き込むため前記プロセッサで実行している適用業務に回答する手段、および前記外部記憶装置から速く離れた場所において前記更新のバックアップコピーを書き込むための手段、から成るシステムである。

【0018】その方法と手段は、ホスト・プロセッサでの以下のステップを含む：

(1) 外部記憶装置における各書き込み操作に回答して、独自の順序番号と外部記憶装置のアドレスを含む書き込みトークンを形成すること、(2) 遠隔事業所と外部記憶装置との間にバスを設定し、新しく生成された書き込みトークンのリスト、および同一または以前のメッセージで送信されたトークン・リストの該当更新データを、それぞれのメッセージが選択的に含むメッセージを送信すること、(3) 遠隔事業所において、またプロセッサから受け取ったそれぞれのメッセージに回答して、データをトークンと照合するために保留書き込み待ち行列を維持すること、合致したトークンおよび更新データについてのみ待ち行列の順で遠隔地において、合致したトークンと更新をスケジュールし書き込むこと。

【0019】本発明においては、プロセッサのOSはDASD外部記憶サブシステムに、ある範囲 (トラック

およびシリンダーのセット)のデータ更新シャドウ化が実施されようとしていることを通知する、「シャドウ化」という言葉は遠隔地へコピーを伝達することを意味している。ついでDASD外部記憶サブシステムは書き込み活動のためにこの範囲を監視して、書き込み操作が進捗していることを順序づけサービスに通知する。この通知は書き込みをされつつあるDASDのトラックとレコードを記述するDASDトークンによってなされる。順序づけサービスは、DASDが作ったトークンを、協同稼働しているシステムのセット内でシャドウ化されようとするすべての他のオペレーションに関するその書き込みオペレーションの時間順を示す順序番号と結合させる。この順序番号が割り当てられそして書き込みデータがサブシステムへ転送された時、サブシステムは入出力書き込み操作を開始したホストへ操作完了を知らせる。トークンと順序番号は共に、指定されたDASDが更新されるという事前情報を提供する非同期メッセージによって、二次事業所へ送信される。

【0020】続いて、データ転送操作は、順序番号とトークンを与えられた後、サブシステムから更新データを検索し、そして順序番号およびトークンと共にその更新データを遠隔事業所または二次事業所へ送信させる。遠隔事業所または二次事業所ではDASD更新データを受信順に待ち行列に並ばせる。その後遠隔事業所は、主事業所でデータが現れたのと同じ順序でシャドウ(バックアップ)・データ・コピーを更新するため、そのローカル記憶情報のアクセス・オペレーションをスケジュールする。主事業所において災害が発生した場合は、二次データ保管場所における回復プロセスは、遠隔事業所または二次事業所で受信した最後の更新データの時刻におけるそのデータの時間的整合性あるイメージを提出するため二次事業所で受信した順次トークンと更新データを解釈することができる。

【0021】前記の目的を満足するためには、遠隔事業所または二次事業所での回復プロセスは、主システム・コンプレックスが、装置の破損はないが、完全に全体的な故障中断を被ったときに主事業所においてDASDに記憶されていた筈のデータと同等なDASDデータのイメージを回復場所において提供できることが必要である。遠隔事業所における回復プロセスの実行に続いて、主事業所において幾分早期に行われた初期システム始動IPLに続いて現れていた筈のデータと全く同様な二次データがシステム・プロセッサで実行される適用業務に現れる筈である。この「早期に」という時間量は、主事業所から回復事業所への距離、主事業所と回復事業所間で利用できる転送チャンネルの帯域幅、およびデータ・ムーバのための利用できる処理能力等の相関要素で変わりうる「データ欠落のウインドウ」である。

【0022】

【実施例】本発明は、システムに含まれる各中央演算処

理装置(CPU)が、IBM社のMVSオペレーティング・システムを持つIBM社のシステム/360またはシステム/370のアーキテクチャーを採用したCPUで構成されるシステムにおいて都合良く実施できる。IBM社のシステム/360アーキテクチャーを採用したCPUに関しては、Amdahl 其他による「データ処理システム」に関する米国特許 3,400,371、1968年9月3日付与、に詳述されている。外部記憶装置を共用アクセスする複数CPUを含むシステム構成に関しては、Luiz 其他による「複数CPUと共用装置アクセス・システムにおける経路非依存の装置予約と再接続」に関する米国特許 4,207,609、1980年6月10日付与、に記述されている。Luiz は仮想コンピューターすなわち「チャンネル」管理と、チャンネルから発せられ記憶サブシステムにより受信され解釈される順次または連鎖のチャンネル・コマンド・ワード(CCW)および同様なコマンドによる外部記憶装置の制御についても記述している。

【0023】MVSオペレーティング・システムもIBM社の出版物、G28-1150、「MVS/拡張アーキテクチャー・システム・プログラミング・ライブラリー」、ボリューム1に記述されている。ローカル・ロック管理、割込みまたは監視によるサブシステム呼出し、タスクの通知と待機などの、標準MVSの詳細およびその他のオペレーティング・システム・サービスの詳細は省略されている。これらのOSサービスは当業者十分に知られているものと信じられる。

【0024】ここで、図1を参照するに、システム、装置のいずれかの障害を起こしがちなシステムにおいて、複写によりデータを保存することにかかわる多様な要因を概念化して説明されている。保護定義域は図示するようにシステムと環境を包含し、装置や媒体のレベルでの可用性の改善に限定される。

【0025】図1に示されているように、一時点複写は歴史的にみて適用業務を中断するものであった。それはバックアップの複写オペレーションが完了するまで実行を延期することを必要とした。現在のシステムはDASD記憶サブシステムからテープへのスケジュールされた転送を行う。一方、先に言及したように、Mikkelsenの審査継続中の出願は適用業務実行の中断を排除し、一時点バックアップをその場の都合に合わせてスケジュールすることを可能にするのではあるが、バックアップする事業所または貯蔵所におけるそのデータの状態は、主事業所における現在のデータ状態より常に遅れている。

【0026】図1に示されているように、通常、リアルタイム複写は、適用業務の実行を中断せず、データの欠落が最小限または皆無であり、そして遠隔事業所を主事業所と同じ最新状態に維持する。DASD群を含み、バックアップ複写することからデータ・ブロックの特別に区分された冗長コード化グループを装置へ非同期式で書込むことにまで及ぶ、前記リアルタイム複写方法は、装

10

20

30

40

50



置と媒体の障害に関してはデータの可用性を拡張するだけであり、主事業所が利用不可能になった場合を考慮していない。

【0027】遠隔事業所への通信に基づく他のリアルタイム複写の方法は、専用バスを必要とし、プロセッサ、記憶サブシステム、および遠隔事業所を完全に閉鎖することになる可能性が高い。これの典型的なものはレーショナル・データ・ベース、自動化銀行システム等のログ・ベースのトランザクション管理システムである。このようなシステムでは、適用業務だけが、コピーの起源と、障害後のデータ状態をトランザクションと一致した状態に回復すべきことの必要性を認識している。

【0028】次に図2を参照するに、複数台のDASDサブシステム5に共用アクセスする1台以上のシステム3と遠隔または二次プロセッサ設置事業所9を連結するインターフェース7で構成される主事業所1が図示されている。本発明のために、これらのシステムの1台は「選定主システム」に指定される。選定主システムはオペレーティング・システム(OS)をもつプロセッサを含んでいる。また、OSはデータ直列化ムーバ(DSM)13と仮想記憶通信アクセス方式(VTAM)11、またはインタフェース7と遠隔事業所9間の通信を達成させるための同様なものを含む。

【0029】DSMの直列化機能部分は各書き込み作業に書き込み順トークンを割り当て、VTAM11が二次事業所(遠隔事業所9)の受信側システム(VTAM15およびデータ・ムーバ17)へ送信するためのメッセージにこれらのトークンを入れる。また、DSM13のデータ・ムーバ機能部分は主DASDに書込まれた変更済みデータ・レコードを受け取り、それらのデータ・レコードとトークンをVTAM11が二次事業所9へ送信するためのメッセージの形に形成する。VTAM11はシステム・ネットワーク体系(SNA)のホスト・プロセッサ部分である。SNAはIBMシステム・ジャーナル、1979年18巻2号に掲載されているP.E.Green、「ネットワーク体系とプロトコル入門」に記述されている。VTAMのオペレーションと実施に関する追加の詳細については、Harris 他による「透過サービス・アクセス機構の適用業務を遠隔ソースへ相互接続するための装置と方法」に関する米国特許4,914,619、1990年4月3日付与、およびAstride 他による「エミュレートした入出力制御装置および共通コマンド変換テーブルと装置アドレス・テーブルを包含する複数プロトコル入出力通信制御装置」に関する米国特許4,855,905、1989年8月8日付与、を参照することができる。VTAM11とVTAM15は機能的には主事業所1と二次事業所9のいずれでも作動できることに留意すべきである。

【0030】さらに図2を参照するに、VTAM11、VTAM15、および主事業所1と二次システム9とのSNAまたはESCON接続7の目的は、その間でメッセー

ジの通信をすることである。これらのメッセージは、シャドウ化(バックアップ)データ・エクステントの確立、進行中の更新を表示するトークン、および二次側データ・コピーに適用される変更済みデータをもつトークンを伝送する。VTAMは主システムと二次システムを結合する単一または複数のチャンネル間接続の広帯域通信ネットワーク上で作動できる。T1回線、T3回線、またはその他の通信サービスは、ESCONや他のチャンネル間直接接続がサポートしうる距離よりも長距離をサポートするために使用される。二次システム9は主事業所1から任意の距離であってよい。

【0031】同様に、二次システム9のプロセッサ部分に常駐のOSは、上に述べたVTAM15のOSベースのサービスに加えてデータ・ムーバの構成要素17をもっている。二次システムのデータ・ムーバ17は、VTAM15を介して主システムから送信されるメッセージを受信し、順序付けトークンを抽出、維持し、そして更新データを二次事業所にあるDASDコピーに適用する。

【0032】バックアップオペレーションは、Mikkelsenの審査継続中の出願に記述されているようなT-ゼロ(T0)・コピー・セッションを確立するための方法と同様に主システムによって、選定されたDASD、ボリュームまたはエクステントに対して確立される。Mikkelsenの方法によれば、主システムのデータ・ムーバは識別されたデータ・エクステントのT0所定時コピーを読取り、そしてバックアップデータの最初の二次コピーを確立するため、そのデータを二次側へ送信する。ひとたびバックアップセッションが開始されると(タイム・ゼロ・コピーの付随の有無にかかわらず)、DASDサブシステムは、データの書き込みのためそれらのエクステントを監視し、書き込みオペレーションが開始されると以下の作業を行う。

【0033】主事業所の外部記憶サブシステムにおける書き込みオペレーション

次に図3を参照するに、図2と同じ構成において、バスに付した番号による順次作業を表示している。このことについては、主事業所ではプロセッサとOSのチャンネル部分は、バックアップDASDアドレスへのデータ書き込み順序を開始するCCWを生成する。これは主事業所DASDサブシステム5(ECKD位置指定レコードCCWに関しては、これはその順序に対する数個のCCWの最初のCCWであろう)によって処理され、そしてデータは、書き込みをするシステムからバス(1)を通過して主事業所のDASDへ転送される。主事業所のDASDサブシステムはバス(2)を通過してDSMの直列化機構13(それはDASDへデータを書き込んでいるシステムと同じシステムに含まれる機構とは限らない)へ書き込みトークンを送信する。要求トークンはバックアップセッション、装置、書き込みトラックとレコードを指定す

る。また要求トークンは固有のサブシステム事象トークンをもつことができる。DSMの直列化機構13は、そのDASDトークンに大域順序番号を割り当て、順序とDASDとの複合トークンを次のメッセージへ付加する。そのメッセージは、二次事業所または遠隔事業所9へ送信するためVTAM11へ渡され、バス(4)を通過して事業所9へ送信される。

【0034】DASDデータ書き込みを実行しているDASDシステム5は、そのデータ書き込みオペレーションを完了すると、バス(3)を通過して要求したシステムへ書き込み完了の信号を送る。(もしこれがホスト・チャンネル・プログラムの終了であれば、書き込み完了は要求プログラムへ通知され、そうでなければチャンネルは、次のオペレーションでチャンネル・プログラムを継続するものと、その信号を解釈する。)最初の書き込みオペレーションに非同期の別個のオペレーションとして、またバス(5)を通過して記述されるDASD供給トークンからのDASDアドレス情報を用いて、ブロック13に常駐のデータ・ムーバーは、DASDサブシステム5からの変更データを読み取る。ついで、データ・ムーバー13は、そのデータをその順序トークンと共に、次のメッセージへ付加し、メッセージは、VTAM11、インターフェース、VTAM15、データ・ムーバー17、主事業所・二次事業所間のバス(6)を包含する回路を通過して二次事業所へ送信される。(書き込み通信量をこなすため複数のデータ・ムーバー・プロセスがあってもよい。)

#### 肯定リターン・トークン

本発明のより強力な実施例は、DSM13が割り当てられた大域順序番号と共に書き込みトークンをDASDサブシステム5へ返送することを伴う。続いてDASDサブシステムは、変更データがデータ・ムーバーへ送信される時に、DSM供給トークンを変更データに付加する。これは同一のレコードに急速に複数の更新がなされる場合の潜在的な不確実性を排除する。このことは前記の基本設計がプログラミングによって適合しなければならぬ条件である。この肯定リターン・トークンによる方法は、DASDサブシステムが保留トークンと付随するデータ(ポインターによる)の持続的な交互の入出力オペレーションの作業リストを維持することを必要とする。これはある種のトラック・キャッシュDASDサブシステム(DASDレコードによる書き込み更新バックアップ)に適合しない設計である。

【0035】二次事業所におけるデータ複写オペレーション

バックアップセッションがひとたび確立されると、二次事業所9は、保留書き込み通知と、バックアップコピーに保存しているDASD19への更新された書き込みデータを受信する。VTAM15は主事業所1からのメッセージを受信し、それらを二次事業所のデータ・ムーバー17へ渡す。これらのメッセージはそれぞれ、主事業所側

データ・ムーバーによって作られ二次側データ・ムーバーにより解釈される、三つのコンテンツ・セグメントから成り立っている。

【0036】次に図4を参照するに、各メッセージの三つのセグメントが図示されている。それらはM0、M1、およびM2として示されている。

【0037】M0はデータ・ムーバー13からデータ・ムーバー17へを見出しで、メッセージ・ストリームの論理連続性を維持するのに役立ち、またメッセージ・コンテンツを識別する。メッセージ・コンテンツは、トークン・データ伝送(M1-M2)を含み、バックアップセッションを確立または終了し、一時点複写の初期コピーまたは他の形式の初期データ・コピーを伝送し、二次側から主事業所側へ論理肯定応答(ACK)および例外通知と回復処置のため通信することができる。M1は、先行するメッセージの後で、一次側DSMによって割り当てられたトークンのセットを含んでいる。これは二次側にデータが存在しない場合の書き込みオペレーションが進行中であることを表す。

【0038】M2はフィールドのセットを含み、それぞれのフィールドはトークンとそれに付随する書き込み(更新)データを含む。

【0039】メッセージの長さや伝送の頻度は、システム間伝送の能率と、進行中の書き込みオペレーションについて二次側への通知に対する遅延時間とのバランスで決まる設計考慮事項である。複数の保留トークンとデータとをメッセージとして一括し、伝送のオーバーヘッドを多数のM1およびM2メッセージ・エレメントに振り分けることにより、能率は向上される。より短い、そしてより頻繁に送信されるメッセージは、より大きな合計伝送オーバーヘッドを生じることになるが、二次側を一次側とより近く同期(遅れが小さい)した状態に保つ働きがある。この概念は、メッセージ・バッファが満杯になった時か、あるいは先行するメッセージが送信された後の時間間隔が満了した時かのいずれか早い方で、一次側データ・ムーバーはメッセージをVTAMへ渡すということにある。メッセージ・バッファのサイズおよび時間間隔は共に選択できるものとする。

【0040】各トークンは、二次側データ・ムーバーがどの物理DASDレコードが更新されるかを判別し、またこれらの書き込みを順序トークンが指定されたのと同じ順序(主事業所側DASDへの書き込みオペレーションを行った適用業務から見た順序と同じである)で配列するのに十分な情報を含んでいる。二次側データ・ムーバーは、トークンがM1セグメントに受信されるときに、まず最初に各トークンを見る。データ・ムーバーは、これらのトークンを保留書き込み待ち行列を維持するために使用する。各トークンに対するデータがM2セグメントに受信されるとき、データは保留中の書き込み待ち行列に含まれているそのデータに対するトークンと照合される

(図3および図4に示されているように)。

【0041】二次側データ・ムーバー17は、二次事業所DASD記憶装置19への書込みオペレーションを、保留中の書込み待ち行列入力の順でスケジュールする。これはバス(7)を通る待機転送として図3に示されている。言い換えるならば、先行する待ち行列エレメントがDASDへの書込みをスケジュールし終わるまでは、所与の待ち行列記入事項に関するデータは書込みをスケジュールされない。データは、それが主事業所側データ・ムーバーによって与えられた時点の関数としてM2メッセージ・セグメントへ到着する。これらのデータの到着は、厳密な更新順にはならないだろう。かくして、活動の任意の時点に、保留中の書込み待ち行列は、完全な記入事項のトークンと付属の書込みデータ、書込みデータがない不完全トークン、これに続くさらに多くの完全トークンと不完全トークンという一連のシーケンスをもつことができる。このことは図5に図示されている。

【0042】主事業所側の障害時の二次側でのDASDデータの回復：上述のメッセージは、バックアップされる主事業所側のデータの回復のため二次側システム9において、情報を提供する。もし主事業所1が、二次側システムでのデータ回復が求められるような災害を被ったら、システムのエレベーション・プロシージャは、バックアップされるデータの全部または一部のために二次側における回復オペレーションを呼び込むことができる。

【0043】次に図5を参照するに、第二事業所を使用するシステムに対する回復プロセスは、保留中の書込み待ち行列から、待ち行列中の最初の不完全トークンまでの保留中のデータ書込みを完了することによって、開始される。これに関連して言えば、データは主事業所におけるいくらか早い時点、すなわち主事業所の障害よりもいくらか以前の時点の所まで整合性を持っている。これは図5の「A」で示されている。このデータは障害時点において、図5の「A」「B」間の記入事項によって表される保留中の書込み待ち行列中の記入事項の残余分と、そして主事業所側でスケジュールされていて実行されていただろうが、それに対してDSMが割り当てた順序トークンが、二次側で図5の後続位置「B」に示されているM1セグメントに受信されなかった書込みオペレーション分だけ、主事業所側より遅れることになる。

【0044】もしシステム回復が一時点まで整合性のあるデータを使用できるならば、それは直ちに回復することができる。これは、主事業所側でシステムのハードウェアやDASDのデータに損壊のない故障が発生した場合(すなわち進行中のジョブと非持久記憶機構のデータは失われてしまい、オープン・データ・セットは整合性のない状態になるだろうが、DASDサブシステムに書き込まれたデータはアクセス可能である)、主事業所側

でのシステムの電源立ち上げおよび初期プログラム・ロード(IPL)のプロセスとほとんど同じであると期待される。

【0045】オペレーションの回復・再開時に、主事業所側で何のデータが欠落したかに関してより多くの情報を必要とするような適用業務に対しては、障害前に主事業所側で進行中の更新が入っていたDASDのトラックおよびレコード、ならびにオペレーションの順序(それが回復に際して必要ならば)を識別するために、保留書込み待ち行列コンテンツが解釈され得る。一部の更新に対しては、先行するいくつかの更新が受信されていなかったため、これらの更新が遠隔DASDに反映されていないにもかかわらず、更新されたデータでさえも遠隔事業所で問い合わせができる。これらの更新は図5の「A」「B」間の順番号131および135として示されている。

【0046】さらに、バックアップセッションを確立することの一部として送信された情報、あるいは協同動作している記憶装置管理システム間で別個に送信された情報は、保留書込みのDASDアドレスをデータセット(保留書込みはこれの一部である)に関連づけるために用いられる。識別されていなかったことになる唯一の進行中の書込みは、主事業所側で作成されたが送信されなかったM1中の書込みおよび障害発生時に伝送中であったM1メッセージである。従って、データセットを整合性ある状態に復元し、システム・オペレーションを再開するために、必要ならば、回復プロセスは、正確にどのデータが欠落したかを障害発生に極めて近い時点まで識別することができる。

【0047】この新方法は、必要なすべてのDASDにわたる更新順序の保存を可能にし、そして主事業所側の書込み性能への影響を最小化する(DASD入出力書込みオペレーションの応答時間の増加を最小化する)記憶装置ベースのデータの遠距離バックアップを行うための手段を定義する。

【0048】従前の既知の方法は次のいずれかであった：(1)データ・ベース・トランザクション・コミット点の認識のような、記憶装置ベースのデータのもつ知識以上のものを要求する、(2)データの整合性を保ったまま、バックアップされたすべてのDASDデータにわたるデータ更新順序の保存ができない、(3)DASD入出力書込みオペレーションと同期して作動するので、入出力オペレーションの応答時間を容認できない程度にまで増加する。

【0049】拡張

本発明の方法と手段に対する一つの拡張は、その組み合わせに関する優れた柔軟性を利用することである。実例として、バックアップセッションは複数の二次事業所で確立することができ、何らかの主事業所側データを一ヶ所の二次事業所へ、また何かを他の事業所へ、そして何

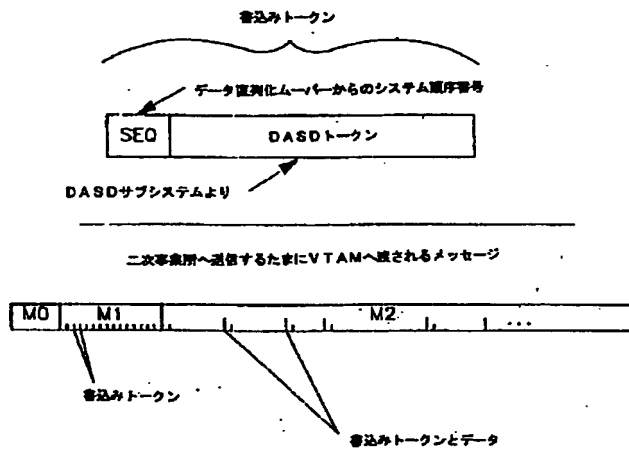
かを二ヶ所以上の事業所へ送信することができる。  
 【0050】請求項に列記されている精神と範囲からから逸脱することなしに、本発明のその他のあらゆる拡張をすることが可能である。

【0051】

【発明の効果】本発明によれば、送信中の更新や、更新コピー発生源が中断された時に受信しなかった更新は別として、非同期的に維持される遠隔バックアップコピー・データの整合性を一次側コピーのそれと同一に維持することができる。また本発明によれば、中断があった場合は遠隔事業所で更新の完全性についての相違や欠落を完全に特定できるような方法と手段が提供される。また、非同期でかつ独立して生成されたコピーと、主事業所から一ヶ所以上の遠隔データバックアップを行う事業所へ与えられる情報とによって、データ・セットの順序と完全性が回復される。

\*

【図4】



\*【図面の簡単な説明】

【図1】複写によってデータを保存する際に考慮すべき多様な要因を示す概念図。

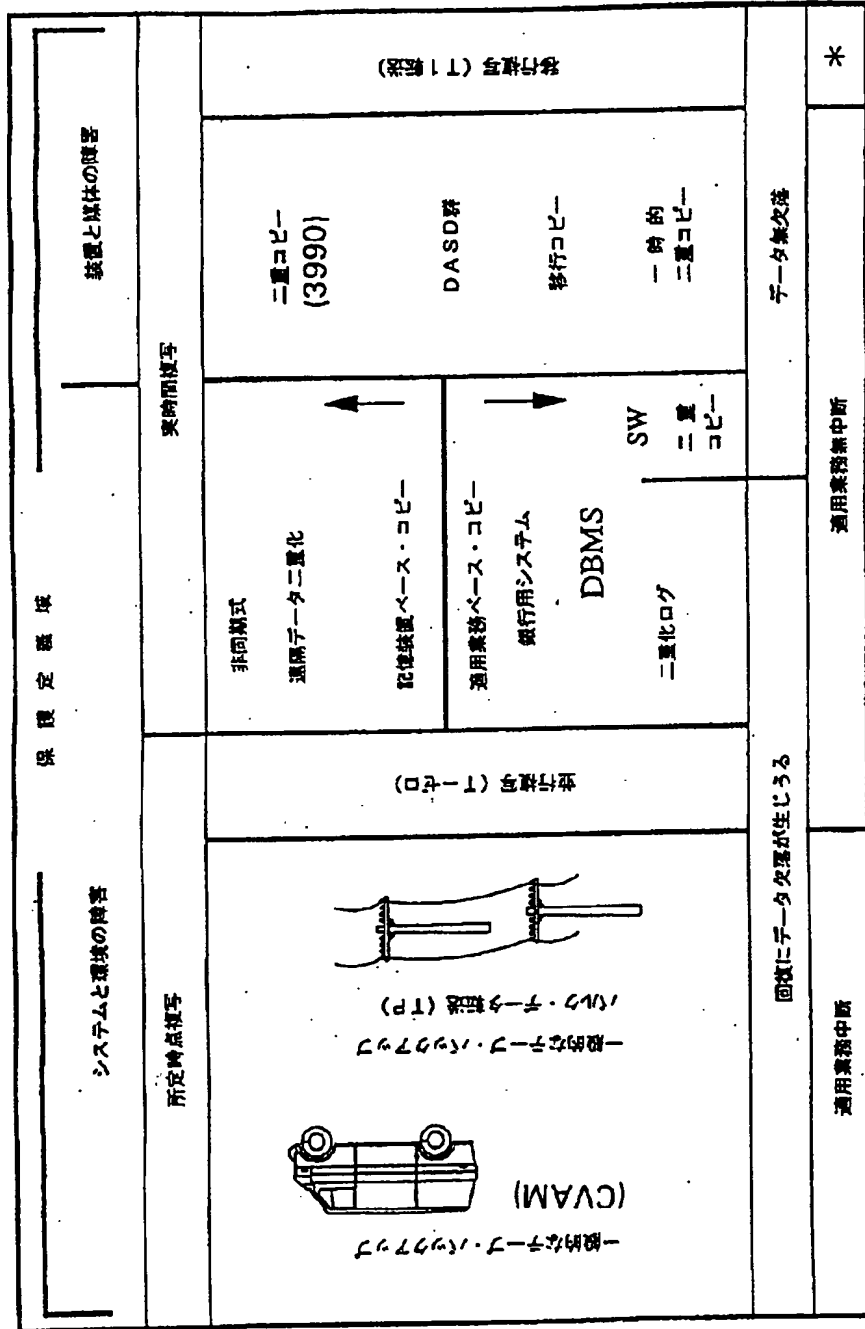
【図2】ホスト・プロセッサ、ホストに接続された外部記憶装置、データがバックアップされる遠隔事業所、および外部記憶サブシステム/遠隔事業所通信インターフェース相互間におけるネットワークの構成図。

【図3】活動順に番号付けをしたパスを表示した図2と同じ構成のネットワーク構成図。

10 【図4】トークンのフォーマット、順番号、およびサブシステムと遠隔事業所間で使用され通信されるメッセージのフォーマットを示す図。

【図5】主事業所における直前の書き込み更新活動のスナップショットとしての遠隔事業所におけるメッセージとトークンの待ち行列を示す図。

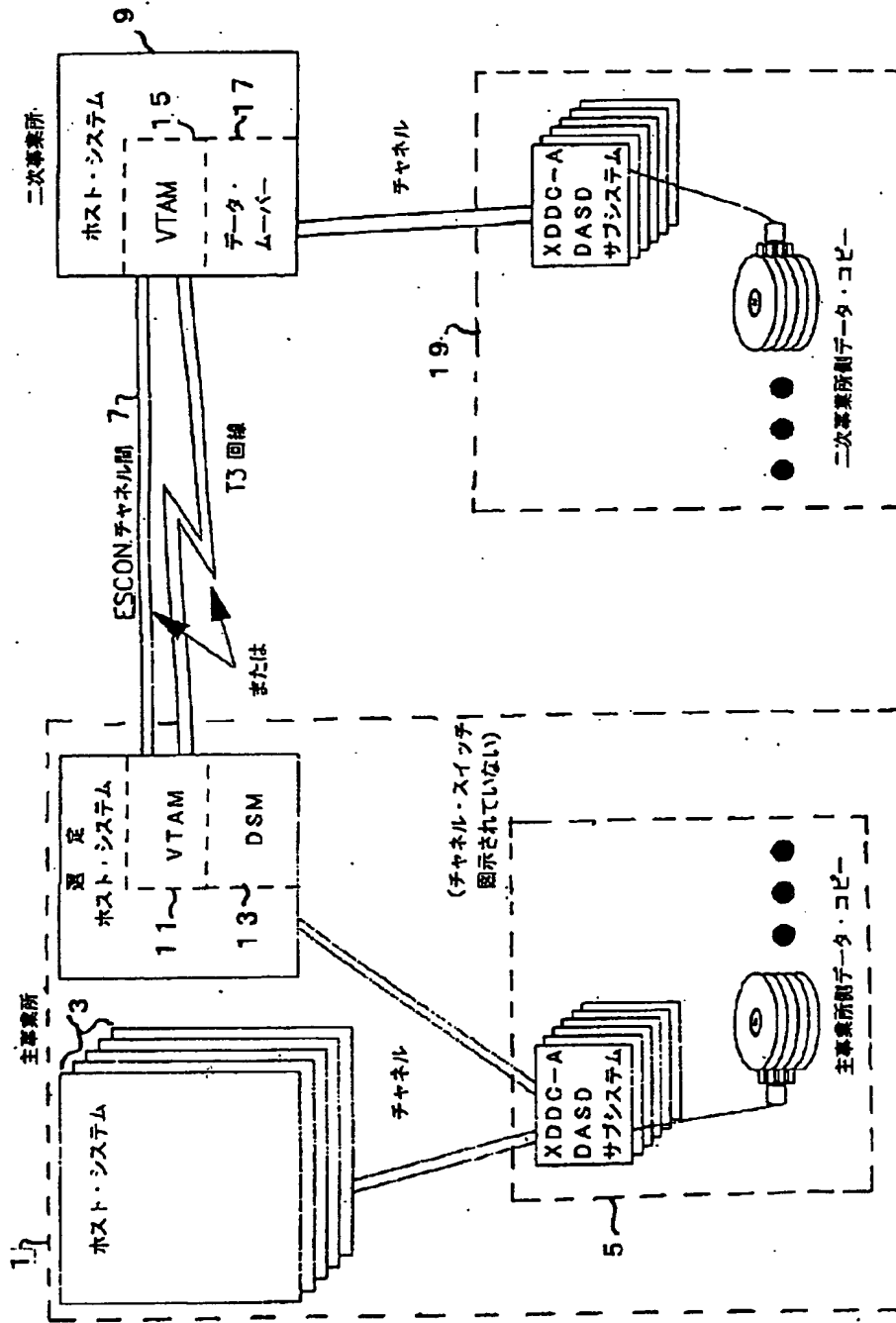
複写によるデータの保存



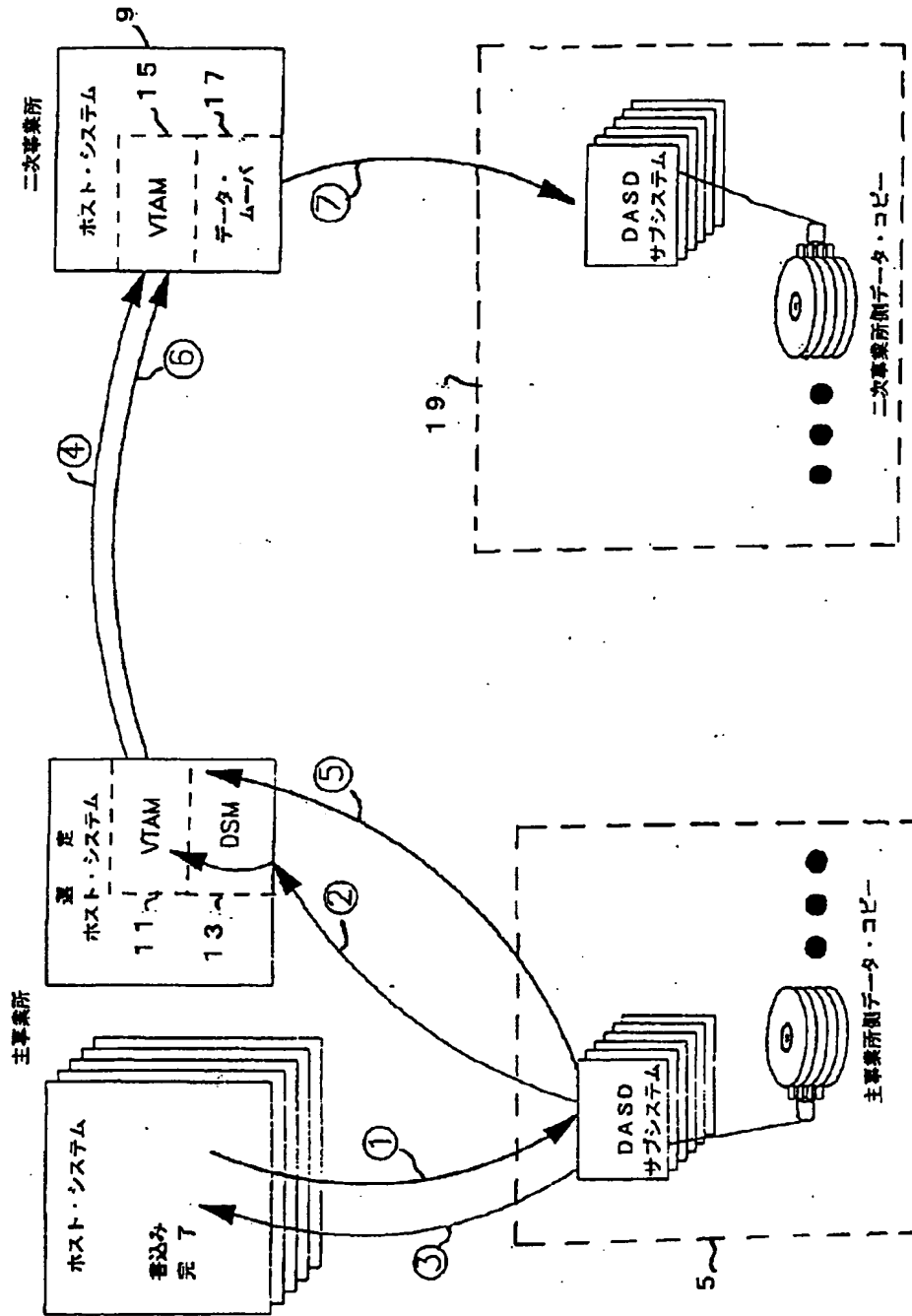
【図1】

(11)

【図2】

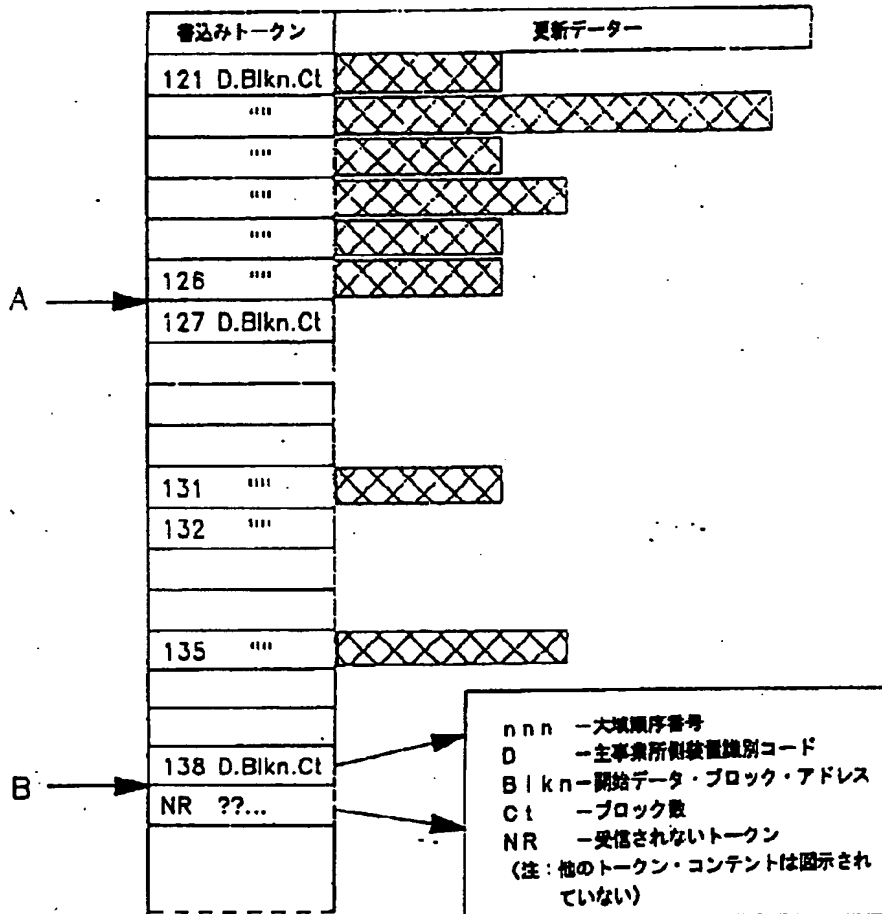


【図3】



【図5】

二次事業所側における保留書込み待ち行列



フロントページの続き

(72)発明者 ジェームズ・イー・マッキルベイン  
アメリカ合衆国カリフォルニア州、サン・  
ホゼ、シルバ・オーク・コート 1118番地